

A statistical method for genetic mapping of sterility genes that exhibit epistasis in remote hybridization of plants using molecular markers in an F_2 population

XIE ShangQian & CHEN JianGuo*

Faculty of Life Sciences, Hubei University, Wuhan 430062, China

Received February 27, 2012; accepted April 20, 2012; published online May 23, 2012

Estimation of the position and effect of sterility genes is an important problem to be solved to understand the sterility mechanism in remote hybridization in plants. In this study, a maximum likelihood (ML) method was used for estimation of the position and effect of sterility genes that exhibit epistasis in an F_2 population using the distorted segregation of markers. The ML solutions for recombination fraction and viability were obtained via an expectation-maximization algorithm. The results of Monte Carlo simulations showed that the estimates of recombination fraction and viability were consistent with their true values. The bias and standard deviation of parameters indicated that a larger sample size, closer linkage and lower viability of sterile genes led to better estimates of the parameters involved. A subset of marker data of the F_2 population derived from a single cross between the rice *japonica* cultivar Nipponbare and the *indica* cultivar Kasalath was analyzed using this method. Eight sterility genes were identified on chromosomes 1, 3, 6, 8 and 10, and significant epistasis was detected among four pairs of sterility genes.

remote hybridization, sterility gene, epistasis, molecular marker, maximum likelihood

Citation: Xie S Q, Chen J G. A statistical method for genetic mapping of sterility genes that exhibit epistasis in remote hybridization of plants using molecular markers in an F_2 population. *Chin Sci Bull*, 2012, 57: 2681–2687, doi: 10.1007/s11434-012-5227-7

Distorted segregation of molecular markers is a phenomenon in which the genotypic frequency of a marker deviates from the expected Mendelian ratio, which is very common in mapping populations derived from remote hybridization in plants. In line-crossing experiments, distorted segregation is usually observed for markers in a number of species such as rice, soybean, cotton and maize [1–5]. One or more segregation distortion loci (SDL) are responsible for the phenomenon. The presence of only a small number of SDL can cause the entire chromosome to distort from Mendelian segregation [6]. These loci are subject to gametic selection, zygotic selection or both [7–9], and most studies show that gametic selection is strongly related to segregation distortion. In maize, four of the identified segregation distortion regions were located in regions where known gametophytic factors occur [3]. Liu et al. [4] constructed soybean F_2 and

recombinant inbred line populations and found that the main reason for distorted segregation was gametic selection. In an F_2 segregation population of *japonica* \times *indica* rice, Guo et al. [10] found 6 segregation distortion regions in which the gametophytic selection genes or sterility genes were located.

Gametophytic selection genes or sterility genes lead to the partial abortion or lethality of female or male gametes, and cause the markers linked to sterility genes to segregate in a distorted manner. The degree of distortion depends on the effect of the sterility genes on male or female gametes and the linkage intensity of the sterility genes with their nearby markers. Hence the information obtained from the distorted markers can be used to estimate the position and effect of the sterility genes. In previous studies on the mapping of sterility genes, Lorieux et al. [11,12] treated the markers as the sterility genes directly, but this is not the usual approach. The estimates obtained in this way will not be correct when sterility genes locate nearby markers. To

*Corresponding author (email: chenjg@yahoo.cn)

improve the accuracy of estimation, methods similar to that of QTL mapping were developed [13–15]. These methods provide a better estimate of the position and effect of a single sterile gene or viability locus. Vogl and Xu [16] developed a Bayesian method to map multipoint viability loci. In addition, Zhu et al. [17,18] estimated the selection coefficient of viability loci and simultaneously reconstructed the genetic linkage map from the adjusted recombination rate in backcross and F_2 populations. However, those methods did not consider the possibility of epistasis between vitality loci or sterility genes. In the present study, we propose a new method that can be used to map sterility genes with epistasis in remote hybridization of plants by using cosegregation data for markers in an F_2 population.

1 Materials and methods

1.1 Genetic model

Consider a mating of $m_1m_1s_1s_1-m_2m_2s_2s_2$ (P_1) \times $M_1M_1S_1S_1-M_2M_2S_2S_2$ (P_2), where M_1-m_1 and M_2-m_2 are two unlinked codominant molecular marker loci, S_1-s_1 and S_2-s_2 are two unlinked sterility loci, S_1-s_1 and M_1-m_1 are assumed to be located on one chromosome, and S_2-s_2 and M_2-m_2 are assumed to be located on a different chromosome. The recombination fraction between M_1 and S_1 is r_1 , and that between M_2 and S_2 is r_2 , and no chiasma interference is assumed. Male gametes produced by the F_1 hybrid are subject to the selection of the sterility genes s_1 and s_2 . The differential viability of male gametes with haplotype s_1S_2 , S_1s_2 and s_1s_2 are expressed as u , v and x relative to that of normal gametes with haplotype S_1S_2 , where x is the epistatic effect between the two sterility genes.

The F_1 hybrid produces 81 kinds of offspring in the F_2 population. The genotype of sterility loci S_1-s_1 and S_2-s_2 cannot be observed directly; only segregation data for the nine genotypic classes of molecular marker M_1 and M_2 can be obtained. Table 1 lists the expected frequencies of the nine kinds of molecular marker genotypes in the F_2 population.

1.2 Estimates of parameters

Let n_i ($i=1,2,\dots,9$) be the observed counts of phenotypes with respect to M_1 and M_2 and n be the total counts of individuals in the F_2 population. The likelihood can be expressed as

$$L = \frac{n!}{\prod_i n_i!} \prod_i p_i^{n_i}, \quad (1)$$

where p_i ($i=1,2,\dots,9$) is the expected relative frequency of the i th phenotypic class of molecular markers M_1 and M_2 .

It is difficult to obtain the explicit solution of ML estimates for all parameters in eq. (1). At this stage, we regard the four kinds of genotypes S_1S_2 , s_1S_2 , S_1s_2 and s_1s_2 as missing

Table 1 Expected frequencies of nine genotypes of molecular markers in the F_2 population

Genotype	Expected relative frequency (p_i) ^{a)}	Observed counts (n_i)
$M_1M_1M_2M_2$	$p_1=[(1-r_1)(1-r_2)+r_1(1-r_2)u+(1-r_1)r_2v+r_1r_2x]/D$	n_1
$M_1M_1M_2m_2$	$p_2=[(1-r_1)(1+v)+r_1(u+x)]/D$	n_2
$M_1M_1m_2m_2$	$p_3=[(1-r_1)(1-r_2)v+r_1(1-r_2)x+(1-r_1)r_2+r_1r_2u]/D$	n_3
$M_1m_1M_2M_2$	$p_4=[(1-r_2)(1+u)+r_2(x+v)]/D$	n_4
$M_1m_1M_2m_2$	$p_5=1/4$	n_5
$M_1m_1m_2m_2$	$p_6=[(1-r_2)(v+x)+r_2(1+u)]/D$	n_6
$m_1m_1M_2M_2$	$p_7=[(1-r_1)(1-r_2)u+r_1(1-r_2)+(1-r_1)r_2x+r_1r_2v]/D$	n_7
$m_1m_1M_2m_2$	$p_8=[(1-r_1)(u+x)+r_1(1+v)]/D$	n_8
$m_1m_1m_2m_2$	$p_9=[(1-r_1)(1-r_2)x+r_1(1-r_2)v+(1-r_1)r_2u+r_1r_2]/D$	n_9
Total	1	n

a) $D=4(1+u+v+x)$.

data, which are assumed to be observed in the complete data set. Then the likelihood function including the four kinds of genotypes of sterility genes is

$$L = \frac{n!}{\prod_{i=1}^9 \prod_{j=1}^4 n_{ij}!} \prod_{i=1}^9 \prod_{j=1}^4 p_{ij}^{n_{ij}}, \quad (2)$$

where n_{ij} and p_{ij} are the observed count of phenotype and the expected relative frequency of the j th kind of sterility gene genotype in the i th kind of marker genotype, respectively. The score for the k th parameter is

$$S_{\theta_k} = \frac{\partial \ln L}{\partial \theta_k} = \sum_{i=1}^9 \sum_{j=1}^4 \frac{n_{ij}}{p_{ij}} \cdot \frac{\partial p_{ij}}{\partial \theta_k}, \quad (3)$$

where $\theta_k = r_1, r_2, u, v, x$. The ML estimates of the parameters can be obtained by setting the above scores to zero and solving the equations

$$r_1 = \left[\sum_{i=1}^3 (n_{i2} + n_{i4}) + \sum_{i=7}^9 (n_{i1} + n_{i3}) \right] / \left(n - \sum_{i=4}^6 n_i \right), \quad (4)$$

$$r_2 = \left[\sum_{i=0}^2 (n_{3 \times i + 1, 3} + n_{3 \times i + 1, 4}) + \sum_{i=1}^3 (n_{3 \times i, 1} + n_{3 \times i, 2}) \right] / \left(n - \sum_{i=0}^2 n_{3 \times i + 2} \right), \quad (5)$$

$$u = \sum_{i=1}^9 n_{i2} / \sum_{i=1}^9 n_{i1}, \quad (6)$$

$$v = \sum_{i=1}^9 n_{i3} / \sum_{i=1}^9 n_{i1}, \quad (7)$$

$$x = \sum_{i=1}^9 n_{i4} / \sum_{i=1}^9 n_{i1}. \quad (8)$$

However, n_{ij} in the above equations is missing genotype counts; hence the estimates cannot be obtained directly from these equations. Here we adopted an EM algorithm [19], including the following steps:

(1) initializing the parameters, such as $r_1^{(0)}=r_2^{(0)}=0.25$, $u^{(0)}=v^{(0)}=x^{(0)}=0.5$;

(2) *E*-step: using the expected relative frequencies p_{ij} and p_i and total counts n_i , n_{ij} can be calculated as follows:

$$n_{ij} = \frac{p_{ij}}{p_i} \times n_i \quad (j = 1, 2, 3, 4). \quad (9)$$

(3) *M*-step: updating the estimate of parameters that can be recalculated by eqs. (4)–(8).

The *E*-step and *M*-step are iterated until convergence occurs, i.e., $|\theta^{(k+1)} - \theta^{(k)}| \leq 10^{-6}$.

1.3 Significance test of the estimates

To test the hypothesis that there is no viability selection, we set the null hypothesis to be $H_0: u = 1, v = 1, x = 1$, and the likelihood ratio (LR) test statistic is

$$\lambda = 2[L(u, v, x, r) - L(u = v = x = 1, r)] \sim \chi^2_{df=3},$$

where $L(\theta)$ is the logarithm of a likelihood function. The multiple test procedure proposed by Benjamini and Hochberg [20] is used to control the false discovery rate (FDR) at $\alpha = \alpha^*/N$, where α^* is 0.01 or 0.05 and N is the number of hypotheses tested.

1.4 Monte Carlo simulation

Simulations were carried out based on two chromosomes of

length 180 cM with 10 equally spaced markers and length 140 cM with eight markers, respectively. Two sterile loci were located on separate chromosomes at 105 and 68 cM, respectively, the first at a distance of 5 cM from the 1 chromosome marker 6 (1M6) and the second at a distance of 8 cM from the 2 chromosome marker 4 (2M4). The following three factors were considered in the simulations: the sample size of the mapping population, the intensity of viability, and the recombination fraction. The sample size ranges from 50 to 500, the intensity of viability of the sterility gene ranges from 0 to 1, and the recombination fraction ranges from 0 to 0.5. Each process was replicated 1000 times, and the bias and standard deviations of the 1000 replicates were used to evaluate the performance of each parameter combination. The purpose of the simulation was to demonstrate that the method works well and the test statistic behaves as expected, and to investigate the range of parameter values in which the method works best.

2 Results

2.1 Results of simulation

An example of the results obtained in an analysis of a data set of 300 individuals with the proposed procedure is listed in Table 2. The parameter estimates for the sterility genes located near the markers 1M6 and 2M4 were almost equal to their true values, and the farther the sterility loci were located from the markers, the greater the estimated values deviated from their true values. When the markers were located far enough from the sterility genes, the effect of viability tended to be 1, which implied that the markers are in normal Mendelian segregation.

The estimates for the 5 parameters were used to obtain

Table 2 Estimation of the position and effect parameters in an analysis of 300 individuals^{a)}

Marker combination	r_1	r_2	u	v	x
1M ₁ -2M ₁	0.225	0.204	0.873	0.735	0.535
1M ₂ -2M ₂	0.227	0.180	0.791	0.599	0.359
1M ₃ -2M ₃	0.213	0.139	0.538	0.349	0.222
1M ₄ -2M ₃	0.168	0.143	0.394	0.336	0.173
1M ₅ -2M ₄	0.110	0.082	0.248	0.196	0.092
1M ₆ -2M ₄	0.037	0.085	0.110	0.195	0.057
1M ₆ -2M ₅	0.035	0.098	0.112	0.224	0.061
1M ₇ -2M ₅	0.077	0.099	0.189	0.227	0.078
1M ₈ -2M ₆	0.146	0.159	0.352	0.380	0.159
1M ₉ -2M ₇	0.186	0.192	0.512	0.527	0.284
1M ₁₀ -2M ₈	0.216	0.219	0.717	0.746	0.456
True value	0.050	0.079	0.1	0.2	0.05

a) The parameters r_1 and r_2 are the recombination fraction between markers and the nearby sterility genes on two different chromosomes; u , v and x are the differential viability of male gametes with haplotype s_1S_2 , S_1s_2 and s_1s_2 , respectively. *Kosambi*: $r_1=0.050$ (5 cM); $r_2=0.079$ (8 cM).

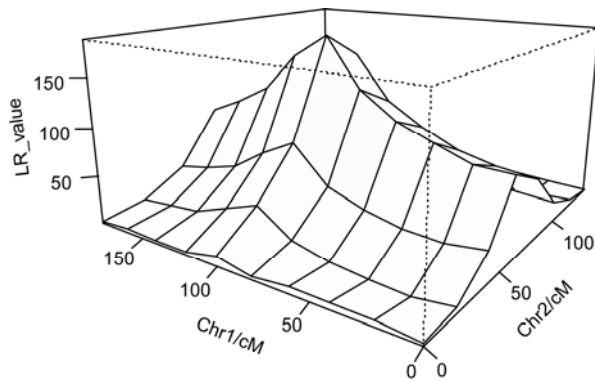


Figure 1 Likelihood ratio (LR) values for the two chromosomes in the hypothesis test.

the LR test statistics (Figure 1). The maximum LR value was obtained at the site of 100 and 60 cM on chromosome 1 and 2, respectively, i.e., the marker combination IM_6-2M_4 , in which the putative sterility genes are located. The cut off α of the multiple test was 6.25×10^{-4} ($\alpha^* = 0.05$) and 1.25×10^{-4} ($\alpha^* = 0.01$), the degrees of freedom was 3, the P -value of a χ^2 -test for the 5 estimates for the IM_6-2M_4 was highly

significant ($P < 1.25 \times 10^{-4}$). This example indicates that the method works well and the test statistic behaves as expected.

The effect of the sample size of the mapping population, the intensity of viability, and recombination fraction on parameter estimation is shown in Figure 2, in which the bias and standard deviation values obtained from 1000 replicates are presented. The bias value decreased with the increase in sample size. When the sample size of the F_2 population exceeded 200, the estimates of the positions and effects of the sterility loci were almost unbiased, and the bias value ultimately was close to 0 (Figure 2(a)), and the standard deviations of the parameter estimates were small (Figure 2(b)). When the recombination fraction r_1 was less than 0.25, the bias of its estimate was relatively small. As the value of r_1 increased, the bias for each parameter increased, among which the bias of u showed the most obvious change; this may be attributed to the fact that u and r_1 were both parameters of the same sterility gene s_1 . The bias of parameter v changed with r_2 in a similar manner. The standard deviation increased with increasing recombination fraction (Figure 2(d)), and the closer the linkage between the molecular marker and sterility gene, the smaller the bias and standard

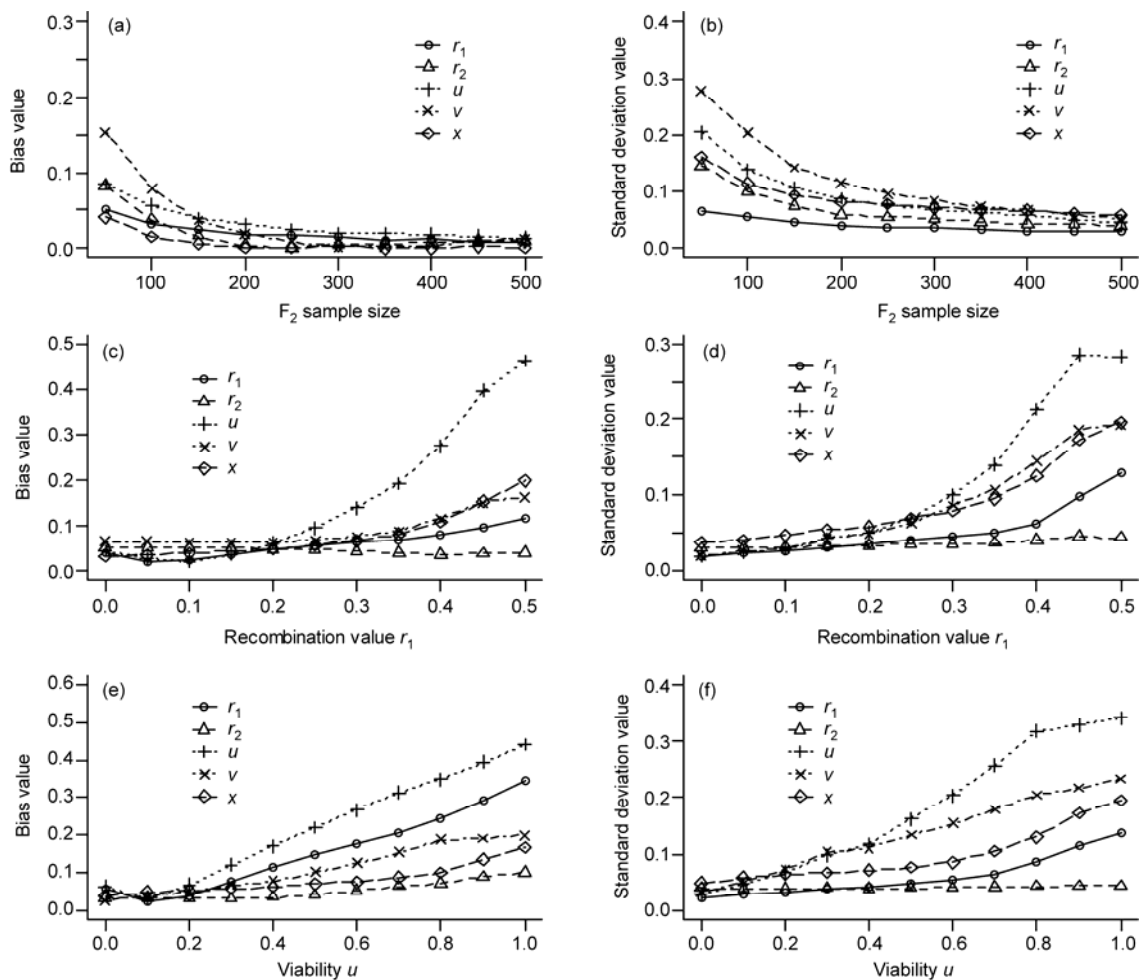


Figure 2 Influence of sample size, recombination fraction and viability on the estimates of parameters.

deviation. When the differential viability u was less than 0.3, the bias and standard deviation of the estimates were small (Figure 2(e) and (f)). The above simulations demonstrated that the proposed method works better with a sample size larger than 200, a strong intensity of viability, and small recombination fraction.

2.2 Real data analysis

A high-density restriction fragment length polymorphism linkage map of rice (*Oryza sativa* L.) covering 1521.6 cM in the Kosambi function was constructed previously using 186 F₂ plants from a single cross between the *japonica* cultivar Nipponbare and the *indica* cultivar Kasalath [21]. According to the results of χ^2 tests with the Joinmap 4.0 software, a subset of markers on chromosomes 1, 3, 6, 8, 9 and 10 de-

viate from Mendelian segregation ratios ($P < 0.05$). This result indicates that these chromosomes may contain sterility genes or selection loci. To save on the computing time required to scan the whole genome, we analyzed 79 markers from these segregation distortion regions and located 8 sterility genes on chromosomes 1, 3, 6, 8 and 10, which were designated as s_1 , s_3 -1, s_3 -2, s_3 -3, s_6 -1, s_6 -2, s_8 , and s_{10} . Significant epistasis was detected among four pairs of the sterility genes (Figure 3). The χ^2 value of markers that are linked with the sterility genes and the position and effect of these sterility genes are listed in Table 3. The loci on chromosomes 3 and 6 were linked with the markers C1186 and L688, respectively. In addition, the loci s_3 -1, s_3 -2, and s_3 -3, and s_6 -1 and s_6 -2, were located close to each other on chromosomes 3 and 6, respectively (Figure 3), and might represent single sterility genes on chromosome 3 and chromosome 6.

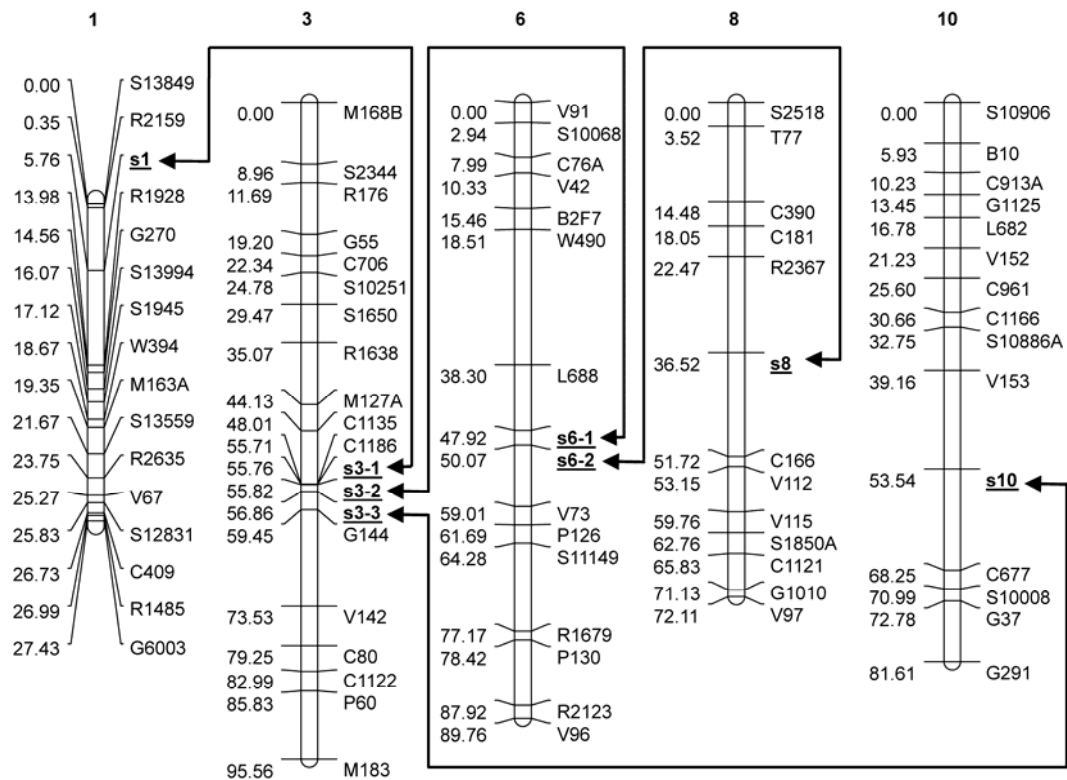


Figure 3 Position and epistasis of sterility genes in the linkage map derived from the F₂ population of Nipponbare × Kasalath.

Table 3 Position and effect of sterility genes detected in the F₂ population of Nipponbare × Kasalath

Interaction ^{a)}				r_1	r_2	u	v	x	$P^b)$
Chr-marker	χ^2	Chr-marker	χ^2						
1-R2159	12.56	3-C1186	68.84	0.054	6.92×10^{-5}	0.528	0.124	5.64×10^{-7}	5.31×10^{-5}
3-C1186	68.84	6-L688	15.85	6.72×10^{-4}	0.125	5.33×10^{-3}	0.307	0.120	3.66×10^{-5}
3-C1186	68.84	10-V153	12.06	0.011	0.140	0.054	0.293	0.036	7.07×10^{-7}
6-L688	15.85	8-R2367	13.22	0.095	0.137	0.105	0.769	0.711	4.53×10^{-2}

a) χ^2 , Chi-squared test value of marker; b) P , the P -value corresponding to the likelihood ratio value of the parameters.

3 Discussion

Sterility genes have been reported in remote hybridization of many plant species. Genetic mapping of this kind of genes is very important theoretically and practically [22–24]. Wang et al. [25] and Yan et al. [26] found that the sterility loci interacted with each other and suggested that is one of the important causes of sterility. On the basis of former research, this paper presents a statistical method that can be used to map sterility genes that exhibit epistasis in remote hybridization of plants using molecular markers in an F_2 population derived from two inbred lines.

Data for molecular markers that show segregation distortion were exploited in the present study to overcome the problem that the genotypes of sterility genes could not be observed directly. Chromosomal regions that cause distorted segregation ratios have been referred to as segregation distortion loci (SDL) [16]. The segregation distortion of markers is caused by numerous factors, such as environmental variables, non-homologous recombination, gene diversion, transposable elements, and transgenic silence, but markers for segregation distortion under these situations do not show a clustered distribution. Clusters of segregation distortion markers usually result from the selective effect of SDL. The different times of action of SDL can be categorized into pollen abortion, pollen tube competition, competitive fertilization, and zygotic selection. The first three stages are collectively termed gametic selection. Sterility genes are one cause of gametic selection, and other causes cannot be quantified. For convenient study of the phenomenon, we assume that gametic selection is due to the effect of sterility genes. The extent of segregation distortion depends both on the distance between the markers and the sterility genes linked with them, and on the effects of the sterility genes. Hence segregation data for molecular markers can be used to estimate the locations and effects of sterility genes. Several methods have been developed to map SDL [15,16,27,28], but none of these methods take into account the existence of epistasis among sterility loci. The method proposed in this paper addresses this problem and does not require a saturated molecular map for estimation of the position and effect of sterility genes, instead only screening of the marker genotype.

To study epistasis among sterility genes, a single effect of a sterility gene can cause the death of an individual, and when more than two sterility genes exist in a gamete, the expression of one gene may influence that of other sterility genes, hence epistasis among sterility genes will complicate the segregation of molecular markers linked to sterility genes and that of other traits linked with them. Therefore, it is improper to analyze epistasis by traditional methods. In this paper, we consider epistasis between two sterility genes as a single effect, which may include many kinds of epistasis in the conventional definition of epistasis, such as complementary effect, additive effect, and duplicate effect. This

approach overcomes practical problems and also reduces the complexity of the method. In the genetic model, we consider the situation that two sterility genes are located on different linkage groups. If two sterility genes exist on the same linkage group, the recombination fraction (designated as r) of the two markers closest to the two genes is known, thus the nine kinds of genotype frequency in the F_2 population are $(1-r)^2/4$, $r(1-r)/2$, $r^2/4$, $r(1-r)/2$, $(1-r)^2/2+r^2/2$, $r(1-r)/2$, $r^2/4$, $r(1-r)/2$, and $(1-r)^2/4$, and can be taken to be the prior probability (f_i). On the basis of the computation formula of the conditional probability, the expected relative

frequency is $f_i p_i / \sum_{j=1}^9 f_j p_j$, where p_i is the unlinked situation frequency in Table 1. Our method is then utilized to estimate the parameters.

Gamete selection can be caused by sterility genes expressed in female or male gametes. We hypothesize that the sterility genes act in the male gametes, which leads to male gamete selection. The proposed method can also be used in the situation in which segregation distortion is to the result of female gamete selection. For species that are difficult to hybridize, F_2 data rather than backcross data are more frequently used for mapping. Where there is definite evidence for the existence of male or female sterility genes, and an F_2 population can be obtained more easily than a backcross population, such as in soybean [29], the proposed method may be useful. When the sterility genes act in both female and male gametes, such as the mutant gene *st8* in soybean [30], this method cannot distinguish the cause of segregation distortion and hence the effect of the sterility genes. Backcross populations are more suitable to discriminate explicitly the cause of segregation distortion. We have developed methods for the mapping of sterility genes that exhibit epistasis in remote hybridization in plants using molecular markers with a backcross population and these will be published soon.

This work was supported by the Natural Science Foundation of Hubei Province (2008CDB082).

- 1 Harushima Y, Kurata N, Yano M, et al. Detection of segregation distortions in an *indica-japonica* rice cross using a high-resolution molecular map. *Theor Appl Genet*, 1996, 92: 145–150
- 2 Li W, Lin Z X, Zhang X L. A novel segregation distortion in intra-specific population of Asian cotton (*Gossypium arboreum* L.) detected by molecular markers. *J Genet Genomics*, 2007, 34: 634–640
- 3 Yan J B, Tang H, Huang Y Q, et al. Genetic analysis of segregation distortion of molecular markers in maize F_2 population (in Chinese). *Acta Genet Sin*, 2003, 30: 913–918
- 4 Liu F, Wu X L, Chen S Y. Segregation distortion of molecular markers in recombinant inbred populations in soybean (*G. max*) (in Chinese). *Acta Genet Sin*, 2000, 27: 883–887
- 5 Lu H, Romer-severson J, Bemardo R. Chromosomal regions associated with segregation distortion in maize. *Theor Appl Genet*, 2002, 105: 622–628
- 6 Xu S. Quantitative trait locus mapping can benefit from segregation distortion. *Genetics*, 2008, 180: 2201–2208

- 7 Faris J D, Laddomada B, Gill B S. Molecular mapping of segregation distortion loci in *Aegilops tauschii*. *Genetics*, 1998, 49: 319–327
- 8 Kärkkäinen K, Koski V, Savolainen O. Geographical variation in the inbreeding depression of Scots pine. *Evolution*, 1996, 50: 111–119
- 9 Kreike D D, Stiekema W J. Reduced recombination and distorted segregation in a *Solanum tuberosum* (2x) × *S. spegazzinii* (2x) hybrid. *Genome*, 1997, 40: 180–187
- 10 Guo J X, Liu Y G. Distorted segregation of molecular markers and their distribution on chromosomes in an *Indica/Japonica* F₂ population of rice (*Oryza sativa* L.) (in Chinese). *Mol Plant Breed*, 2003, 1: 331–336
- 11 Lorieux M, Goffinet B, Perrier X, et al. Maximum-likelihood models for mapping genetic markers showing segregation distortion. 1. Backcross population. *Theor Appl Genet*, 1995, 90: 73–80
- 12 Lorieux M, Perrier X, Goffinet B, et al. Maximum-likelihood models for mapping genetic markers showing segregation distortion. 2. F₂ populations. *Theor Appl Genet*, 1995, 90: 81–89
- 13 Fu Y B, Ritland K. On estimating the linkage of marker genes to viability genes controlling inbreeding depression. *Theor Appl Genet*, 1994, 88: 925–932
- 14 Mitchell-Olds T. Interval mapping of viability loci causing heterosis in abidopsis. *Genetics*, 1995, 140: 1105–1109
- 15 Luo L, Zhang Y M, Xu S Z. A quantitative genetics model for viability selection. *Heredity*, 2005, 94: 347–355
- 16 Vogl C, Xu S Z. Multipoint mapping of viability of segregation distorting loci using molecular markers. *Genetics*, 2000, 155: 1439–1447
- 17 Zhu C, Wang C S, Zhang Y M. Modeling segregation distortion for viability selection I. Reconstruction of genetic linkage maps with distorted markers. *Theor Appl Genet*, 2007, 114: 295–305
- 18 Zhu C, Wang F, Wang J, et al. Reconstruction of linkage maps in the distorted segregation populations of backcross, doubled haploid and recombinant inbred lines. *Chin Sci Bull*, 2007, 52: 1648–1653
- 19 Dempster A P, Laird N M, Rubin D B. Maximum likelihood from incomplete data via EM algorithm. *J Royal Stat Soc Ser B*, 1977, 39: 1–8
- 20 Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J Royal Stat Soc Ser C Appl Stat*, 1995, 57: 289–300
- 21 Harushima Y, Yano M, Shomura A, et al. A high-density rice genetic linkage map with 2275 markers using a single F₂ population. *Genetics*, 1998, 148: 479–494
- 22 Lin S Y, Ikehashi H. A gamete abortion locus detected by segregation distortion of isozyme locus Est-9 in wide crosses of rice (*Oryza sativa* L.). *Euphytica*, 1993, 67: 35–40
- 23 Liu Y S, Sun J S, Zhou K D. Cytological basis causing spikelet sterility of intersubspecific hybrid in *Oryza sativa* (in Chinese). *J Mol Cell Biol*, 1997, 30: 335–341
- 24 Chen J G, Walsh B. Method for the mapping of a female partial-sterile locus on a molecular marker linkage map. *Theor Appl Genet*, 2009, 119: 1085–1091
- 25 Wang J, Liu K, Xu G, et al. The high level of wide-compatibility of variety ‘Dular’ has a complex genetic basis. *Theor Appl Genet*, 1998, 97: 407–412
- 26 Yan C J, Liang G H, Gu S L, et al. Molecular marker analysis and genetic basis for sterility of typical *Indica/Japonica* hybrids (in Chinese). *Acta Genet Sin*, 2003, 3: 76–85
- 27 Wang C, Zhu C, Zhai H, et al. Mapping segregation distortion loci and quantitative trait loci for spikelet sterility in rice (*Oryza sativa* L.). *Genetics Res*, 2005, 86: 97–106
- 28 Luo L, Xu S. Mapping viability loci using molecular markers. *Heredity*, 2003, 90: 459–467
- 29 Pereira T N S, Lersten N R, Palmer R G. Genetic and cytological analyses of a partial-female-sterile mutant (ps-1) in soybean (*Glycine max*; Leguminosae). *Am J Bot*, 1997, 84: 781–791
- 30 Kato K K, Palmer R G. Molecular mapping of the male-sterile, female-sterile mutant gene (*sr8*) in soybean. *J Hered*, 2003, 94: 425–428

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.